# *Mismatch Distributions of mtDNA Reveal Recent Human Population Expansions*

STEPHEN T. SHERRY,[1] ALAN R. ROGERS,[2] HENRY HARPENDING,[1] HIMLA SOODYALL,[3] TREFOR JENKINS,[3] AND MARK STONEKING[1]

*Abstract* Although many genetic studies of human evolution have tried to make distinctions between the replacement and the multiregional evolution hypotheses, current methods and data have not resolved the issue. However, new advances in nucleotide divergence theory can complement these investigations with a description of human demographic behavior during the late Middle and Upper Paleolithic (approximately the last 250,000 years). Restriction fragment length polymorphism (RFLP) and DNA sequence analyses of human mitochondrial DNAs (mtDNAs) from 25 ethnic and racial groups indicate that significant expansions occurred during the late Middle and Upper Paleolithic in 23 of the 25 populations examined. Estimates for the individual group expansion times are consistently less than 100,000 years ago with a mean expansion time of approximately 40,000 years ago. The dramatic expansions suggested by these data occurred well after modern human anatomy appeared, approximately 100,000 years ago, but are concordant with archeological evidence for the expansion of modern human technology, approximately 50,000 years ago.

The haploid maternal inheritance of human mitochondrial DNA (mtDNA) means that any variation observed when analyzing mtDNA types is the simple product of mutational divergence without recombination. From this perspective human mtDNA types can be regarded as the tips of a single tree of evolutionary descent embedded in the species gene pool. We consider mtDNA types to be lineages, and the future fate of any lineage probabilistically depends on the size of the population in which it is carried. Population growth lowers the probability of extinction of each lineage and embeds a starlike burst of lineages in the genealogical history of the population. New lineages start their passage through the

[1]Department of Anthropology, Pennsylvania State University, University Park, PA 16802.
[2]Department of Anthropology, University of Utah, Salt Lake City, UT 84102.
[3]MRC Human Ecogenetics Research Unit, Department of Human Genetics, South African Institute for Medical Research and University of Witwatersrand, Johannesburg, South Africa.

*Human Biology*, October 1994, v. 66, no. 5, pp. 761–775.


KEY WORDS: HUMAN EVOLUTION, MITOCHONDRIAL DNA, UPPER PALEOLITHIC, DEMOGRAPHY, MISMATCH DISTRIBUTION

population as identical copies of preexpansion lineages. As a result, they establish a common time interval for most later coalescent events (Slatkin and Hudson 1991).

This identity between pairs of alleles is reflected as a mode at zero in the mismatch distribution (Rogers and Harpending 1992; Harpending et al. 1993). As time progresses and as the mutational process separates once-identical lineages from one another, the mode at zero in the mismatch distribution advances through increasing values of $i$, the average number of nucleotide differences between sequence pairs. In an infinite sites model this value increases at the rate $2u$ for haploid sequence data, where $u$ is the sum of per-nucleotide mutation rates in the region of DNA under study (Rogers and Harpending 1992). With knowledge of the mutation rate $u$ we can map these demographic events from units of mutational time into units of real time.

The theory developed by Li (1977) and extended by Rogers and Harpending (1992) to describe the behavior of the distribution of pairwise differences in a nonequilibrium population suggests that pairwise difference distributions can provide estimates for Middle Paleolithic female effective population size and dates of rapid growth. Subsequent simulations of the model of sudden expansion have shown that parameters can be estimated with narrow confidence intervals and that computational estimation is much faster than with alternative tree-based methods (Felsenstein 1992).

Here, we describe the effects of sample size, technique of data acquisition, and designation of population (ethnic resolution) on the behavior of the two parameters in our sudden expansion model of rapid human population growth. The parameters are

$$\theta = 2Nu, \tag{1}$$

the conventional product of the preexpansion female effective population size $N$ and the locus-specific mutation rate $u$ (per generation); and

$$\tau = 2tu, \tag{2}$$

the date of expansion measured in units of mutational time, where $t$ is time in generations. A generation length of 25 years was used for unit conversion. The populations described here were sampled for mtDNA variation and assayed with techniques developed to detect either restriction fragment length polymorphisms (RFLPs) (Cann et al. 1987) or sequence differences (Vigilant et al. 1989).

## Materials and Methods

**Geographic Origins of mtDNA Samples.** The sample characteristics of 25 populations are described in Table 1. Five populations were pre-

**Table 1.** mtDNA Sample Characteristics

| *Population* | *N* | *Data Type* | | | *Reference* |
|---|---|---|---|---|---|
| | | *RFLP* | *HVS-I* | *HVS-I/II* | |
| 1 Australian | 21 | X | | | Cann et al. (1987) |
| 2 African 1 | 20 | X | | | Cann et al. (1987) |
| 3 African 2 | 8 | | X | X | Vigilant et al. (1989) |
| 4 African 3 | 10 | | X | | Horai and Hayasaka (1990) |
| 5 East Asian 1 | 23 | | X | X | Vigilant et al. (1989, 1991) |
| 6 East Asian 2 | 34 | X | | | Cann et al. (1987) |
| 7 East Asian 3 | 10 | | X | | Horai and Hayasaka (1990) |
| 8 Middle East | 42 | | X | | DiRienzo and Wilson (1991) |
| 9 Bantu speakers | 36 | | X | X | Soodyall (1993) |
| 10 Dama | 20 | | X | X | Soodyall (1993) |
| 11 European 1 | 19 | | X | X | Vigilant et al. (1989, 1991) |
| 12 European 2 | 47 | X | | | Cann et al. (1987) |
| 13 European 3 | 20 | | X | | Horai and Hayasaka (1990) |
| 14 Hadza | 17 | | X | X | Vigilant et al. (1991) |
| 15 Herero | 38 | | X | X | Vigilant et al. (1989, 1991); Soodyall (1993) |
| 16 Japanese | 61 | | X | | Horai and Hayasaka (1990) |
| 17 !Kung | 64 | | X | X | Vigilant et al. (1989, 1991); Soodyall (1993) |
| 18 Nama | 17 | | X | X | Soodyall (1993) |
| 19 Nuu Chah Nulth | 63 | | X | | Ward et al. (1991) |
| 20 Highland Papua New Guinea | 29 | | X | X | Vigilant et al. (1991); Stoneking et al. (1992) |
| 21 Papua New Guinea | 119 | X | | | Stoneking et al. (1990) |
| 22 Mbuti Pygmy | 19 | | X | X | Vigilant et al. (1989) |
| 23 Biaka Pygmy | 17 | | X | X | Vigilant et al. (1989) |
| 24 Sardinian | 69 | | X | | DiRienzo and Wilson (1991) |
| 25 Yoruban | 14 | | X | X | Vigilant et al. (1989, 1991) |

viously characterized for mtDNA variation with high-resolution RFLP maps (Cann et al. 1987; Stoneking et al. 1990), whereas 20 populations were previously characterized by nucleotide sequences of polymerase chain reaction (PCR) amplified or cloned products of hypervariable control region segments (Vigilant et al. 1989, 1991; DiRienzo and Wilson 1991; Horai and Hayasaka 1990; Ward et al. 1991; Stoneking et al. 1992; Soodyall 1993). Some studies (DiRienzo and Wilson 1991; Horai and Hayasaka 1990; Ward et al. 1991) determined 360 bp of the first hypervariable segment (HVS-I) only, whereas others (Vigilant et al. 1989, 1991; Stoneking et al. 1992; Soodyall 1993) determined an additional 393 bp of the second hypervariable segment (HVS-II).

**Construction of Pairwise Difference Distributions.** Pairwise difference distributions were computed for all populations reported in Table 1. Populations that had been sequenced at both HVS-I and HVS-II were analyzed twice, once using only HVS-I for comparison with the exclusively HVS-I samples and again using the additional 393 bp of HVS-II to test the stability of the distributions with the addition of more information. All 572 sequences were aligned with the human reference sequence (Anderson et al. 1981) and insertion-deletion gaps and positions with missing data were removed before analysis. This resulted in 195 common sites (out of 360) over 572 individuals for HVS-I and 428 common sites (out of 753) over 317 individuals for the combined HVS-I and HVS-II.

An alternative procedure for handling missing data was explored in which population parameters were estimated using the maximum number of common sites in each ethnic sample. Because missing sites varied in both number and position from sample to sample, parameter estimates and standard errors were not strictly comparable across ethnic groups and hence were normalized to a common number of sites (i.e., the maximum number under the ideal conditions of no missing data). The data cited by Harpending et al. (1993) were based on this treatment of missing data and were normalized over 360 sites for HVS-I and 753 sites for HVS-I and HVS-II (data not shown). Our mean expansion times therefore differ slightly from the values reported by Harpending et al. (1993). The average difference in expansion times between the two procedures was 9,800 years for HVS-I and HVS-II and 18,200 years for HVS-I; the time spans of the confidence intervals were almost identical. Because mean population expansion times estimated by either procedure consistently fell in the early Upper Paleolithic (except for the recently bottlenecked populations), our conclusions remain unchanged.

**Computing Estimators.** The mean $m$ and variance $v$ were computed for each sample and the parameters $\theta$ and $\tau$ (defined earlier) were estimated by the method of moments in which

$$\hat{\theta} = (v - m)^{1/2}, \tag{3}$$

$$\hat{\tau} = m - \hat{\theta} \tag{4}$$

(Rogers 1994). The standard errors reported for values of $\hat{\tau}$ were estimated by simulation with the coalescent algorithm described by Harpending et al. (1993). Coalescent trees with $N$ tips and population-specific values for $\hat{\theta}$ and $\hat{\tau}$ were randomly generated 1000 times for each population analyzed. The standard deviation was computed for each collection of 1000 trees and used as an estimate for the standard error of $\hat{\tau}$.

An approximate 95% confidence interval for $\hat{T}$ (mean expansion

time in years) was empirically computed to account for the error in estimates of $\tau$ and $u$, where $u = m_T\mu$, the quantity $m_T$ is the number of nucleotides under study, and $\mu$ is the standard mtDNA evolutionary rate (Stoneking et al. 1992). Since most of the sites eliminated because of missing data were flanking ends of sequences with little internal variation, the estimated divergence rates used here ($0.41 \pm 0.06$ per site per million years for HVS-I and $0.25 \pm 0.03$ per site per million years for HVS-I and HVS-II) are higher than previously reported estimates (Stoneking et al. 1992; Ward et al. 1991). Our estimates of $\mathrm{SE}(\hat{u})$ come from Stoneking et al. (1992) and unpublished data (for HVS-I and HVS-II).

Given the relationship between $\tau$ and time defined as

$$T = \tau/2u \tag{5}$$

(Rogers and Harpending 1992), the lower bound $\hat{T}_{\mathrm{LB}}$ and the upper bound $\hat{T}_{\mathrm{UB}}$ of the confidence interval for $T$ were calculated as

$$\hat{T}_{\mathrm{LB}} = \frac{\hat{\tau} - 2\mathrm{SE}(\hat{\tau})}{2[\hat{u} + 2\mathrm{SE}(\hat{u})]}, \tag{6}$$

$$\hat{T}_{\mathrm{UB}} = \frac{\hat{\tau} + 2\mathrm{SE}(\hat{\tau})}{2[\hat{u} - 2\mathrm{SE}(\hat{u})]}. \tag{7}$$

A consequence of the model used to estimate $\tau$ is that the variance of the distribution must exceed the mean. Although this relationship between the variance and the mean often failed with simulated data, it failed for only 3 of the 41 pairwise difference distributions analyzed. In these three cases $\hat{\theta}$ was set to 0 in subsequent computations, which equates $\hat{\tau}$ to the mean sequence divergence (MSD). Although the method of moments also assumes that the population has grown to an infinitely large size, in practice equivalent results are obtained with this model so long as there is at least 100-fold growth (Rogers 1994). This is because with this level of growth the starlike burst of new lineages is virtually assured preservation (and hence statistical detection) in a nonrecombinant gene tree until a later, more dramatic demographic event overwrites it.

The smoothness of the mismatch distributions is characterized by the raggedness statistic $r$, described by Harpending et al. (1993). If there is a maximum of $d$ mutational differences in the mismatch distribution, we define $r$ as

$$r = \sum_{i=1}^{d} (x_i - x_{i-1})^2, \tag{8}$$

where $x_i$ is the frequency of pairs of mtDNA types that differ by exactly $i$ substitutions. Although this statistic is simply an ad hoc measure of high-frequency variation in the distribution, it does differentiate between

mismatch distributions from equilibrium and expanded populations (Harpending et al. 1993; Harpending 1994). Raggedness measures are reported for the empirical distributions in Table 2 and were also computed for simulated data under three demographic scenarios: equilibrium, 100-fold expansion, and 1000-fold expansion. In each scenario 1000 simulations were performed for $\tau = 0, 1, \ldots, 8$ and the mean raggedness value was reported.

Summary values for each data set (RFLP, HVS-I, combined HVS-I and HVS-II) were computed by grouping all individuals in one analysis. These results are labeled "World sample" in Table 2. In a similar fashion regional analyses were performed for Africans, Asians, Europeans, and non-Africans (Asians and Europeans combined) in the HVS-I and combined HVS-I and HVS-II data sets.

## Results

**Effects of Sample Size.** Analysis of simulated data (Figure 1) suggests that the variance of $\hat{\tau}$ is independent of the mean, thus providing us with some degree of confidence in dating ancient population events. In the interest of minimizing this variance most efficiently, we have also investigated the effect of sample size on reducing the standard error in estimates of $\hat{\tau}$ (Figure 2). This analysis suggests that increasing sample size will reduce the variance of $\hat{\tau}$, but this improvement is marginal for sample sizes larger than 50 or so.

**Technique of Data Acquisition.** We were interested in the relative performance of high-resolution RFLP maps compared with DNA sequences. Analysis of high-resolution RFLP maps found evidence of Upper Paleolithic expansions in all five populations examined (Table 2). Mean population expansion times were based on an mtDNA evolutionary rate of 3% per million years (Stoneking et al. 1986). The time ranges for RFLP data reported in Table 2 were calculated from the reported 2% and 4% limits. Because the true 95% confidence interval for the mtDNA divergence rate may be larger than this estimate, the expansion time ranges reported for these RFLP data are not true 95% confidence intervals.

**Sensitivity to Sample Bias.** Because virtually all the populations represented in this study were sampled opportunistically (i.e., subject inclusion based on immediate availability for study) in clinical or field situations, we devised a comparison to demonstrate the reliability of our estimator in sharply contrasted cases of sample composition. In this example two mismatch distributions were constructed from the Papua New Guinea RFLP data (Figure 3). The first was based on the original field

**Table 2.** Parameter Estimates and Population Expansion Times[a]

| *Population* | $N$ | $\hat{\theta}$ | $\hat{\tau} \pm \mathrm{SE}(\hat{\tau})$ | $\hat{T}$ *(ka)* | $\hat{T}_{LB}$–$\hat{T}_{UB}$ | $r$ |
|---|---|---|---|---|---|---|
| Analysis of HVS-I | | | | | | |
| African 2, 3 | 18 | 0.69 | 4.48 ± 0.89 | 56.0 | 26.1–110.7 | 0.027 |
| East Asian 1 | 23 | 0.00 | 4.99 ± 0.64 | 62.4 | 35.9–110.9 | 0.019 |
| Bantu speakers | 36 | 1.12 | 3.91 ± 0.94 | 48.8 | 19.6–102.4 | 0.020 |
| Biaka Pygmy | 17 | 3.16 | 2.29 ± 1.95 | 28.7 | 0.0–109.4 | 0.153 |
| Dama | 20 | 1.52 | 3.12 ± 1.21 | 39.0 | 6.7–97.9 | 0.020 |
| European 1, 3 | 39 | 0.00 | 2.38 ± 0.31 | 29.7 | 17.0–53.0 | 0.045 |
| Hadza | 17 | 1.99 | 0.83 ± 1.51 | 10.3 | 0.0–68.0 | 0.369 |
| Herero | 38 | 1.01 | 0.43 ± 0.75 | 5.4 | 0.0–34.1 | 0.107 |
| Japanese | 61 | 0.22 | 3.52 ± 0.47 | 43.9 | 25.0–78.8 | 0.018 |
| !Kung | 64 | 0.85 | 3.50 ± 0.73 | 43.7 | 19.7–87.7 | 0.012 |
| Mbuti Pygmy | 19 | 2.76 | 1.95 ± 1.76 | 24.4 | 0.0–96.7 | 0.194 |
| Middle East | 42 | 0.00 | 3.76 ± 0.43 | 47.0 | 20.2–81.7 | 0.034 |
| Nama | 17 | 1.30 | 4.06 ± 1.21 | 50.7 | 15.8–114.6 | 0.019 |
| Nuu Chah Nulth | 63 | 0.65 | 2.87 ± 0.61 | 35.9 | 15.9–72.3 | 0.019 |
| Highland Papua New Guinea | 29 | 2.44 | 3.14 ± 1.55 | 39.2 | 0.4–110.3 | 0.011 |
| Sardinian | 69 | 0.74 | 1.19 ± 0.67 | 14.8 | 0.0–44.7 | 0.027 |
| Yoruban | 14 | 0.66 | 3.10 ± 0.89 | 38.7 | 12.8–86.3 | 0.038 |
| World sample | 568 | 0.66 | 4.26 ± 0.56 | 53.3 | 30.4–95.0 | 0.012 |
| Africans | 242 | 0.65 | 4.64 ± 0.57 | 58.1 | 33.9–102.2 | 0.012 |
| Asians | 176 | 0.69 | 4.04 ± 0.56 | 50.5 | 28.3–91.1 | 0.015 |
| Europeans | 150 | 0.27 | 2.35 ± 0.37 | 29.4 | 15.5–92.8 | 0.026 |
| Non-Africans | 326 | 0.70 | 3.27 ± 0.57 | 40.9 | 20.7–77.9 | 0.017 |
| Analysis of HVS-I and HVS-II | | | | | | |
| African 2 | 8 | 4.28 | 7.03 ± 3.13 | 65.7 | 5.8–163.4 | 0.046 |
| East Asian 1 | 23 | 1.37 | 6.11 ± 1.24 | 57.2 | 33.3–115.9 | 0.011 |
| Bantu speakers | 36 | 1.51 | 6.93 ± 1.25 | 64.8 | 33.3–115.9 | 0.007 |
| Biaka Pygmy | 17 | 4.73 | 4.70 ± 2.70 | 44.0 | 0.0–124.2 | 0.031 |
| Dama | 20 | 3.32 | 5.09 ± 2.08 | 47.5 | 7.0–113.7 | 0.017 |
| European 1 | 19 | 1.27 | 3.30 ± 1.09 | 30.8 | 8.4–67.3 | 0.022 |
| Hadza | 17 | 2.97 | 0.96 ± 2.01 | 8.9 | 0.0–61.2 | 0.369 |
| Herero | 38 | 3.03 | 0.04 ± 1.70 | 0.3 | 0.0–42.3 | 0.057 |
| !Kung | 64 | 2.48 | 3.95 ± 1.52 | 36.9 | 6.8–85.9 | 0.008 |
| Mbuti Pygmy | 19 | 3.72 | 3.01 ± 2.19 | 28.2 | 0.0–90.8 | 0.178 |
| Nama | 17 | 2.13 | 5.33 ± 1.62 | 49.7 | 15.7–105.4 | 0.028 |
| Highland Papua New Guinea | 29 | 3.23 | 4.47 ± 1.90 | 41.8 | 5.0–101.6 | 0.006 |
| Yoruban | 14 | 2.19 | 4.78 ± 1.71 | 44.7 | 10.2–100.8 | 0.022 |
| World sample | 313 | 1.60 | 7.39 ± 1.05 | 69.3 | 39.7–117.9 | 0.005 |
| Africans | 242 | 1.86 | 6.86 ± 1.18 | 64.2 | 33.9–113.4 | 0.006 |
| Asians | 52 | 1.97 | 6.36 ± 1.31 | 59.5 | 28.1–110.4 | 0.005 |
| Europeans | 19 | 1.27 | 3.30 ± 1.09 | 30.8 | 8.4–67.4 | 0.022 |
| Non-Africans | 71 | 1.85 | 5.75 ± 1.20 | 53.8 | 25.2–100.2 | 0.005 |
| Analysis of RFLP data | | | | | | |
| African 1 | 20 | 4.07 | 8.95 ± 4.90 | 99.8 | 34.3–225.6 | 0.013 |
| East Asian 2 | 34 | 1.97 | 7.88 ± 3.02 | 87.8 | 41.2–177.5 | 0.007 |
| European 2 | 47 | 1.70 | 4.99 ± 2.44 | 55.6 | 21.6–121.0 | 0.015 |
| Australian | 21 | 2.49 | 4.40 ± 3.41 | 49.0 | 8.3–127.2 | 0.008 |
| Papua New Guinea | 119 | 1.95 | 4.26 ± 2.58 | 47.5 | 14.2–111.5 | 0.010 |
| World sample | 241 | 2.45 | 5.50 ± 3.02 | 61.3 | 21.0–138.3 | 0.012 |

a. The following additional parameters were used in the computations: HVS-I, 195 bp common sequence length, $\hat{u} = 4 \times 10^{-5} \pm 5.85 \times 10^{-6}$ per site per year; HVS-I and HVS-II, 428 bp common sequence length, $\hat{u} = 5.35 \times 10^{-5} \pm 6.42 \times 10^{-6}$ per site per year; RFLP, 236 variable sites, $\hat{u} = 4.5 \times 10^{-5} \pm 7.5 \times 10^{-6}$ per site per year.

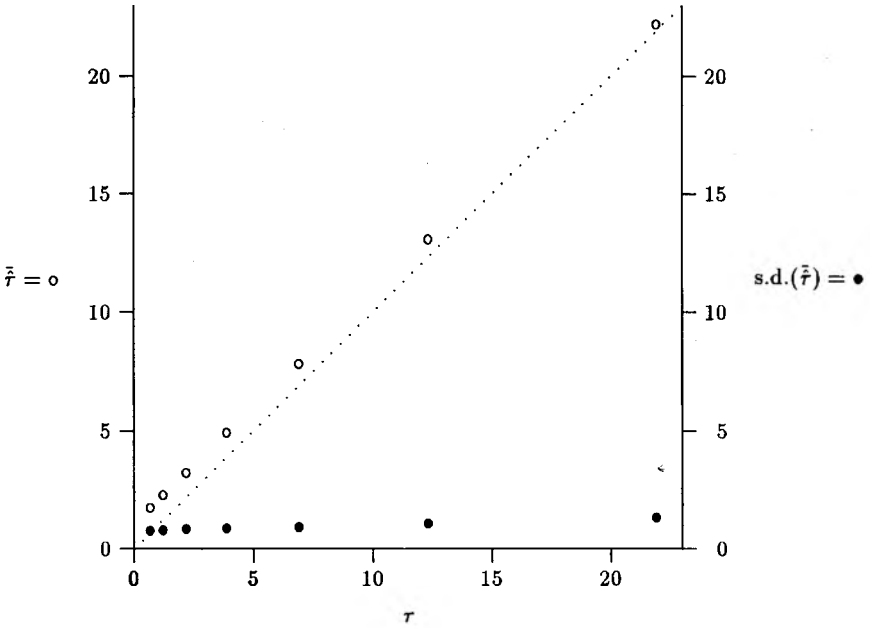


**Figure 1.** Estimation of $\tau$ and SE($\tau$) from simulated data. Each open circle shows the mean estimate of $\tau$ obtained from 1000 simulated sets of data. Each simulation assumed that $\theta_0 = 2.50031$ and $\theta_1 = 2500$ and that 147 individuals were sampled. The filled circles show the standard deviations of these estimates. The true value of $\tau$ is graphed as a dotted line.

collections consisting of 119 individuals from various highland and coastal localities (Stoneking et al. 1990). The second distribution was constructed from 50 RFLP haplotypes that were preselected from the previous set of 119 to maximize Papua New Guinea mtDNA type diversity. Although the sampling schemes are different, the estimates of $\hat{\tau}$ are virtually identical (4.34 for group 1 and 4.57 for group 2).

If "ethnicity" is restricted in meaning to designate local differentiation within a major continental area, this result underscores our observation that the genetic variation used in this analysis is much older than the origins of modern ethnic groups. In other words, ethnic differentiation is a much younger process than the population expansions of the late Pleistocene, and as a result, any African, Asian, or European selected (for the sake of example) would be as good a choice as any other in a study at this level of resolution.

**Human Population Expansions.** Expansion time estimates from the three data sets (RFLP, HVS-I, and combined HVS-I and HVS-II) are presented in Table 2 and summarized in Figure 4. These results suggest

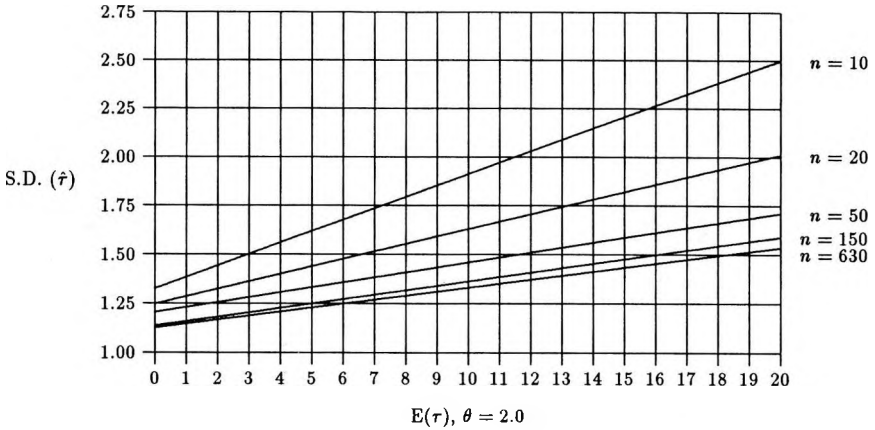


**Figure 2.** Standard deviations of $E(\tau)$ for sample sizes of 10, 20, 50, 150, and 630. For each sample size, 1000 simulations were created for 21 values of $E(\tau) = 0, 1, 2, \ldots, 20$ with $\theta = 2.0$ and an infinite final population size ($1 \times 10^{16}$). The slopes shown in this figure are linear regressions of the simulated means and standard deviations.

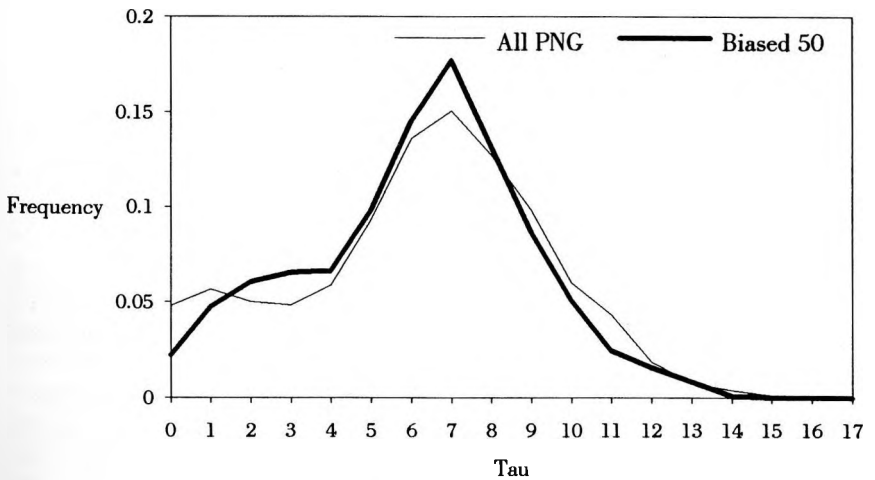


**Figure 3.** Comparison of Papua New Guinea (PNG) mismatch distributions. One distribution was constructed from mtDNA RFLP haplotype data for all 119 individuals, whereas the second is a subset of the first in which 50 individuals were selected for type diversity. The resulting estimates for $\tau$ are virtually identical (all PNG $\hat{\tau} = 4.34$, biased PNG $\hat{\tau} = 4.57$), and map to expansion times of 48,800 years ago and 50,100 years ago using the 3% divergence rate estimate discussed in the text.

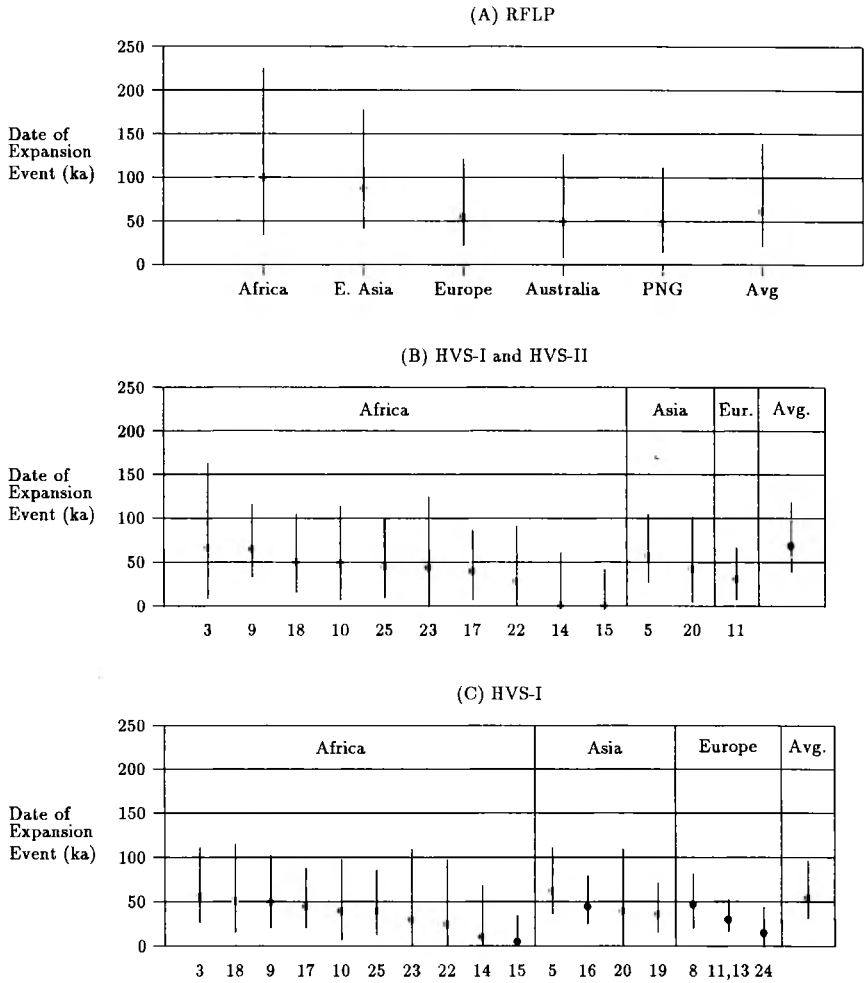


**Figure 4.** Estimated population expansion times. Filled circles indicate estimates of population expansion times, and the vertical rules represent an approximate 95% confidence interval (except for the RFLP data) based on the values reported in Table 2. Populations are numbered as in Table 1 and are grouped by regional affiliation.

that major episodes of human population expansion have been relatively recent events, with the majority occurring 30,000–65,000 years ago. The dates reported are based on estimates of $u$ calibrated for the data set used here and strongly depend on the variation in this single molecule. Rather than emphasizing the mean times of population expansion, which are subject to considerable error, we encourage an inspection of the time ranges spanned by the confidence interval $T_{LB}$–$T_{UB}$.

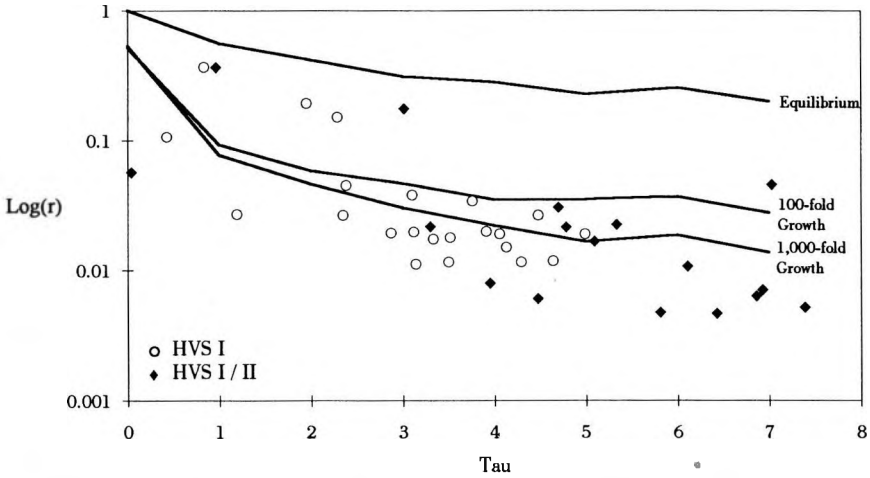


**Figure 5.** Open circles are empirical raggedness measures for the population samples in Table 1. For comparison, populations were simulated under three demographic scenarios (equilibrium, 100-fold growth, 1000-fold growth) over a range of expansion times ($\tau$ = 0, 1, . . ., 8); the average of $r$ for 1000 simulations of each combination is shown. Most of the distributions show a signature of past population growth.

The mean expansion times, averaged over all groups in HVS-I and in the combined HVS-I and HVS-II, are 37,900 years ago and 42,700 years ago, respectively. The two populations with markedly more recent inferred expansion times (Hadza and Herero) are known or suspected to have experienced recent demographic bottlenecks (Pennington and Harpending 1993; Blurton Jones et al. 1992)—events that would account for the low levels of mtDNA variability observed in these samples.

Values of the raggedness statistic (Table 2) can also be evaluated to assess the magnitude of population growth (Figure 5). These data show good agreement between most human populations and scenarios of at least 100-fold growth. In this analysis only the Hadza and Mbuti Pygmy raggedness measures show no evidence of growth. The estimated expansion times are consistent with other African populations; perhaps demographic events have altered the raggedness measures for these two populations.

The Herero are an example of yet a third situation in which estimates of both $\tau$ and $r$ are small. The small estimate of $\tau$ indicates a recent bottleneck event, but the small value of $r$ suggests a more dramatic population recovery for the Herero than the Hadza and Mbuti Pygmies. On the other hand, the regional composite groups of Africans, Europeans, Asians, non-Africans, and the general world sample all show strong evidence for dramatic population growth during the early Upper Paleolithic.

## Discussion

The analysis of mtDNA mismatch distributions reveals evidence of prehistoric population expansions in virtually every population examined. When the data are collapsed into three continental regions (Africa, Asia, Europe), the mean population expansion times span an approximate 35,000-year interval beginning roughly 65,000 years ago and ending about 30,000 years ago (Table 2, regional analyses). Although the African expansion appears to be the earliest, the large confidence intervals preclude any definitive ordering of regional expansion times.

The signatures of population growth during the early Upper Paleolithic reported here suggest a more complex history of human evolution than can be accommodated by either the multiregional or the rapid-replacement hypotheses. It should be noted that these hypotheses are only two of a vast number of scenarios used to model recent human evolution. The profound differences in their proposed descriptions of population movement and evolution easily lead us to accept them as extreme alternatives. Unfortunately, this conceptualization belies the subtle nature of their respective assumptions and the nature of the data that could reject their explicit predictions (Long 1993).

The predictions of the most extreme version of the rapid-replacement model require a single signature of population expansion during the Lower or Middle Paleolithic in all contemporary populations. These signatures were not detected in the present study. The multiregional evolution scenario, on the other hand, cannot be tested as easily with these data, because its dynamics are couched in measures of peripheral gene flow between widely dispersed hominid groups. The best that we can do with these data is estimate the preexpansion population size and compare it with minimal population sizes for a globally dispersed yet genetically admixed population. Our estimates of a global preexpansion effective population size range from 40 to 600 females, which translates into a total population size of less than 3000 at best—less than 2% of conservative estimates for a sufficiently dense *Homo erectus* world population (Harpending et al. 1993).

Alternative models, such as the Weak Garden of Eden scenario favored by Harpending et al. (1993), have been proposed that are consistent with mtDNA mismatch analysis. In this scenario the processes of population separation and expansion are decoupled in time: Population separation begins soon after the appearance of anatomically modern *Homo sapiens* about 80,000–100,000 years ago, whereas episodes of dramatic population expansion are delayed another 50,000 years.

**Implications for Human Evolution.** In the archeological record the time range of 45,000–35,000 years ago serves as an approximate bound-

ary for an apparently dramatic change in human behavior (Knecht et al. 1993). Artifact industries, such as the European Middle Paleolithic (or Mousterian) that existed before 50,000 years ago, are characterized by remarkably little variability through time and space. The number of readily identifiable stone tool types that they contain is relatively small, and formal bone artifacts, decorative items, and art objects are all but absent. In the interval between 50,000 and 40,000 years ago the Mousterian and similar industries were widely replaced by clearly more advanced industries, including those of the Upper Paleolithic. In these lithic assemblages the degree of geographic and temporal variability is far greater, the number of readily recognizable stone artifact types is much larger, and formal bone artifacts and art objects are common (Klein 1989, 1992; Harrold 1989).

However, the transition in morphological skeletal traits to an anatomically modern form appears substantially earlier in the fossil record. Modern human anatomy appears first in Africa and the Near East roughly 100,000 years ago (Bar-Yosef 1992; Stringer and Andrews 1988), and these early specimens are associated with fairly uniform Middle Paleolithic/Middle Stone Age industries (Stiner, personal communication, 1993). Klein (1989, 1992) noted this discrepancy and proposed a scenario for the expansion of anatomically modern humans during the transition from the Middle to the Upper Paleolithic based in large part on a superior material culture and advancements in lithic technology. Our estimates of population expansion times center around 40,000 years ago and thus are in good agreement with the regional appearances of Upper Paleolithic tool traditions. Whether there is a causal connection between these events is much less clear.

An alternative reading of the tool evidence provided by Stiner and Kuhn (personal communication, 1993) is that there was not intensive "diversifying selection" on technology during the Early and Middle Paleolithic. This intensity changed at the transition from the Mousterian to the Upper Paleolithic (in Eurasia, at least). According to this scenario, the expanding variety of tool forms one sees after about 40,000 years ago suggests a gradual change in the role of technology in human adaptations, not a discrete technological advance per se. The net effect of improvements in technical efficiency throughout the Upper Paleolithic would be gradual and possibly only sufficient to buffer populations from crashes during future glacial events (18,000–35,000 years ago).

The population expansions described here should not be confused with previous studies describing the mtDNA common human ancestor or coalescent-based models of replacement by anatomically modern humans (Cann et al. 1987; Vigilant et al. 1989, 1991; Merriwether et al. 1991). Assuming that estimates of $\mu$ are correct to within an order of magnitude, episodes of human population expansion during the Lower

or Middle Paleolithic would not be resolved with the model employed here. This result is not surprising because the data set analyzed here carries no information about population histories before the common human mtDNA ancestor. Before the common ancestor there are no pairs of lineages to compare.

In summary, an analysis of human mtDNA mismatch distributions was performed and signatures of dramatic past population growth were detected. The collection of population expansion times was compared with the archeological and fossil records and was found to be broadly consistent with Klein's (1992) hypothesis of a major technology-based expansion during the Upper Paleolithic. We are currently seeking further resolution of regional population histories by incorporating data from independently evolving nuclear loci into this sudden expansion model.

*Acknowledgments* We are grateful for comments and helpful discussion from J. Crow, J. Hey, S. Horai, R. Klein, N. Takahata, K. Weiss, M. Stiner, S. Khun, and A. Mukherjee. We thank R. Zauhar for computational assistance. This research was supported in part by the National Science Foundation under grant BNS 90-20567 awarded to Mark Stoneking and by the US Department of Health and Human Services under grant MGN 1-R29-GM39593 awarded to Alan Rogers.

*Received 23 August 1993; revision received 31 January 1994.*

## Literature Cited

Anderson, S., A.T. Bankier, B.G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465.

Bar-Yosef, O. 1992. Middle Paleolithic chronology and the transition to the Upper Paleolithic in southwest Asia. In *Continuity or Replacement: Controversies in* Homo sapiens *Evolution*, G. Bräuer and F. Smith, eds. Rotterdam, Netherlands: A.A. Balkema, 261–272.

Blurton Jones, N.G., L.C. Smith, J.F. O'Connell et al. 1992. Demography of the Hadza, an increasing and high density population of savanna foragers. *Am. J. Phys. Anthropol.* 89:159–181.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

DiRienzo, A., and A. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 88:1597–1601.

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: Inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genet. Res.* 59:139–147.

Harpending, H.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* 66(4):591–600.

Harpending, H.C., S.T. Sherry, A.R. Rogers et al. 1993. Genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–496.

Harrold, F. 1989. Mousterian, Chatelperronian, and Early Aurignacian in Western Europe: Continuity or dicontinuity? In *The Human Revolution*, P. Mellars and C. Stringer, eds. Princeton, NJ: Princeton University Press, 677–713.

Horai, S., and K. Hayasaka. 1990. Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* 46:828–842.

Klein, R. 1989. Biological and behavioral perspectives on modern human origins in southern Africa. In *The Human Revolution*, P. Mellars and C. Stringer, eds. Princeton, NJ: Princeton University Press, 529– 546.

Klein, R. 1992. The archeology of modern human origins. *Evol. Anthropol.* 1:5–14.

Knecht, H., A. Pike-Tay, and R. White. 1993. Introduction. In *Before Lascaux: The Complex Record of the Early Upper Paleolithic*, H. Knecht, A. Pike-Tay, and R. White, eds. Boca Raton, FL: CRC Press, 1–4.

Li, W.-H. 1977. Distribution of two randomly chosen cistrons in a finite population. *Genetics* 85:331–337.

Long, J.C. 1993. Human molecular phylogenetics. *Annu. Rev. Anthropol.* 22:251–272.

Merriwether, D.A., A.G. Clark, S.W. Ballinger et al. 1991. The structure of human mitochondrial DNA variation. *J. Molec. Evol.* 33:543–555.

Pennington, R., and H. Harpending. 1993. *The Structure of an African Pastoralist Society: Demography, History and Ecology of the Ngamiland Herero.* Monographs in Human Population Biology Series. Oxford, England: Clarendon Press.

Rogers, A. 1992. Error introduced by the infinite sites model. *Molec. Biol. Evol.* 9:1181–1184.

Rogers, A. 1994. Genetic evidence for a Pleistocene population explosion. *Evolution* (in press).

Rogers, A., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise differences. *Molec. Biol. Evol.* 9:552–569.

Slatkin, M., and R.R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562.

Soodyall, H. 1993. Mitochondrial DNA polymorphisms in southern African populations. Ph.D. thesis, University of Witwatersrand, Johannesburg, South Africa.

Stoneking, M., K. Bhatia, and A.C. Wilson. 1986. Rate of sequence divergence from restriction maps of mitochondrial DNAs from Papua New Guinea. *Cold Spring Harbor Symp. Quant. Biol.* L1:433–439.

Stoneking, M., L.B. Jorde, K. Bhatia et al. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733.

Stoneking, M., S. Sherry, A. Redd et al. 1992. New approaches to dating suggest a recent age for the human mtDNA ancestor. *Philos. Trans. R. Soc. London* 337:167–175.

Stringer, C.B., and P. Andrews. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263–1268.

Vigilant, L., R. Pennington, H. Harpending et al. 1989. Mitochondrial DNA sequences in hairs from a Southern African population. *Proc. Natl. Acad. Sci. USA* 86:9350–9354.

Vigilant, L., M. Stoneking, H. Harpending et al. 1991. African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507.

Ward, R., B. Frazier, K. Dew-Jager et al. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA* 88:8720–8724.